

Coventry University Repository for the Virtual Environment  
(CURVE)

**Author name:** Gardner, S.

**Title:** 'Integrating ethnographic, multidimensional, corpus linguistic and systemic functional approaches to genre description : an illustration through university history and engineering assignments

**Article & version:**

**Original citation & hyperlink:**

Gardner, S. (2008). Integrating ethnographic, multidimensional, corpus linguistic and systemic functional approaches to genre description : an illustration through university history and engineering assignments. In E. Steiner & S. Neumann (Eds). *Proceedings of the 19<sup>th</sup> European Systemic Functional Linguistics Conference and Workshop*. Saarbrücken, Germany.

[urn:nbn:de:bsz:291-scidok-16889](https://nbn-resolving.org/urn:nbn:de:bsz:291-scidok-16889)

<http://scidok.sulb.uni-saarland.de/volltexte/2008/1688/>

This document is made available under a Creative- commons Attribution-NonCommercial-NoDerivs 2.0 Germany licence. For full licence terms see:

<http://creativecommons.org/licenses/by-nc-nd/2.0/de/deed.en>. When citing the document, please use the following identifiers: URN: urn:nbn:de:bsz:291-scidok-16889 URL: <http://scidok.sulb.uni-saarland.de/volltexte/2008/1688/>

Available in the CURVE Research Collection: May 2012

<http://curve.coventry.ac.uk/open>

# INTEGRATING ETHNOGRAPHIC, MULTIDIMENSIONAL, CORPUS LINGUISTIC AND SYSTEMIC FUNCTIONAL APPROACHES TO GENRE DESCRIPTION: AN ILLUSTRATION THROUGH UNIVERSITY HISTORY AND ENGINEERING ASSIGNMENTS

Sheena Gardner, University of Birmingham, UK

Our research aims to describe genres of assessed writing at British universities (ESRC RES-000-23-0800). To this end we have developed a corpus of 2800 texts from four years of study across four broad disciplinary groupings. Our research design integrates a corpus linguistic account of formal features in the corpus with an ethnographic investigation of the disciplinary context, a multi-dimensional analysis of register, and a functional linguistic analysis of genres.

In this paper I illustrate this design with examples from history and engineering. The contextual information shows that history students write mostly essays, written as pedagogical genres, while engineering students engage in a wide range of written assignments: scientific papers are written as if to report findings to an academic audience; funding proposals are written as if to persuade a professional readership; posters are designed to inform a lay audience (e.g. visitors to a transport museum); and reflective journals are written for personal and professional development. The writing process also differs. Some assignments are written individually whereas others involve teamwork.

The multidimensional analysis conducted by Biber in Arizona suggests dimensions along which the registers of History and Engineering differ, and those where they converge. Interpreting this data is not straightforward, and this paper interprets findings from analysis using Biber's original dimensions. Further interpretation is anticipated using the BAWE specific dimensions.

The analysis of assignments into genres identifies generic stages, social purpose, and typical lexico-grammatical features. The analysis of 60 assignments from three years of study written by 17 students in 21 modules on 59 topics shows that university history assignments correspond closely to five genres described in Coffin's study of secondary school history student writing: Analytical Discussion, Analytical Exposition, Factorial Explanation, Consequential Explanation and Challenge. It also explores the extent to which lexico-grammatical features suggest progression.

The analysis of 205 Engineering assignments from three years of study across 15 degree programmes, suggests twelve specific genres, including Laboratory Report, Design Proposal, Product Evaluation, Design Report, Exercise, and Research Report. There is evidence of progression from first to final year assignments in terms of generic structure. This is surprisingly absent in the history assignments analysed.

The paper concludes with discussion of issues and benefits encountered in blending a systemic functional analysis of genre, with a multidimensional analysis of register, corpus linguistic description of the corpus, and an ethnographically-informed account of disciplinary context. Just as an ethnographically informed genre analysis increases its trustworthiness, and corpus linguistic analyses can increase the power of manual genre analyses, so too does genre-informed multidimensional analysis enable deeper interpretation of findings which can then be checked with the discourse community.

KEYWORDS: genre, academic disciplines, corpus, multiple method research design, student writing

## 1 INTRODUCTION

Our Investigation of Genres of Assessed Writing in British Higher Education<sup>1</sup> aims to produce a rich description of genres of assessed student writing through the integration of ethnographic, functional linguistic and multidimensional computational approaches to text description. Specifically, the research design falls into five strands: the development of the corpus, the description of the formal features of the corpus, the investigation of discourse community perspectives, the multidimensional analysis of register, and the systemic functional analysis of genres. While an examination from different perspectives would enhance our understanding of writing at university, we perhaps did not anticipate all the issues that would arise in developing a corpus which could be used in these different ways.

Many of the issues we encountered in corpus development might not have arisen had we not been aiming to blend functional, interpretive and formal approaches to our data. They relate, for instance, to whether we were analysing the assignments as a mass of ‘text,’ as computational, form-focused corpus linguistic methods favour; or whether we were analysing the assignments as individual instances of genres, to be analysed as whole texts with specific meanings interpreted in their educational contexts. Related to this were issues of whether different theoretical perspectives would unnecessarily constrain our analyses, or provide the detail and rigour desired. The difficulties of bringing together corpus linguistics and systemic functional linguistics are indicated in Hunston and Thompson’s tongue in cheek mutual critiques: is computerised corpus analysis “reductive, insufficiently related to the texts of which the corpus is composed or to their social contexts”? Does the Systemic Functional Linguistic (SFL) “view of

---

<sup>1</sup> The project An investigation of genres of assessed writing in British Higher Education, which was funded by the Economic and Social Research Council (project number RES-000-23-0800) from 2004 to 2007, includes the development of the British Academic Written English corpus at the Universities of Warwick, Reading and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for English Language Teacher Education, Warwick), Paul Thompson (Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes).

language close and constrain observation, in that aspects of language that cannot be accounted for in terms of the three-part system are not accounted for at all”? (Thompson and Hunston, 2006:3) Bringing in our ethnographic perspectives, we could add ‘are linguistic approaches too narrowly focused on language without sufficient attention to the understandings and intentions of those who produced it and the purposes it is intended to serve?’

In this paper I discuss the five strands of our research design, highlighting areas where approaches combine or clash, and illustrate their contributions through research on the disciplines of History and Engineering. Section 2 focuses on corpus development; Section 3 on context; Section 4 on multidimensional analysis; Section 5 on more conventional corpus analysis readily afforded by digital corpora; and Section 6 on genre analysis.

## 2 CORPUS DEVELOPMENT

There are no other comparable corpora of student writing as earlier studies of academic writing in more than one discipline have tended to construct small focused corpora (e.g. Charles 2006, Samraj 2005) and to draw on texts in the public domain such as theses and published research articles (e.g. Groom 2005, Hyland 2004). Our aim was that the Corpus should consist of good written assignments from many different disciplines and indeed from across universities. To operationalise this, we assumed that assignments which had been awarded good marks by subject tutors would qualify as well written in the disciplinary communities. Assignments were collected from four British universities, representing different types of university, and together providing a spread of departments across the 19 programme categories used for university admissions.

### 2.1 SAMPLING ISSUES

One issue that arose early on was how we could ensure reasonable representation in the corpus from across disciplines and years of study. We considered different sampling methods from Random Sampling, through Stratified Sampling, Cluster Sampling, and Opportunistic or Convenience Sampling to Purposive or Judgemental Sampling. With an eye on genre analysis, I was very much in favour of purposive sampling; we wanted to have several instances of each text type in order to be able to describe specific genres in disciplines. To do this, we could collect at least five instances of any given assignment type (e.g. ethnography, essay, exercise, report, book review) in any given assignment year. And to be able to see development across years, we could try and collect assignment types across years in similar content areas (e.g. from a physical geography or gender in sociology strand within the broader disciplines of geography and sociology). This could be done by tracking pre-requisites. A balance of core and optional modules might be desirable. And the number of assignments submitted by any given student should be limited. The more we considered how we would ideally target specific assignments, the more we realised that such constraints would work counter to the aim of building a substantial corpus from across the spectrum of university writing.

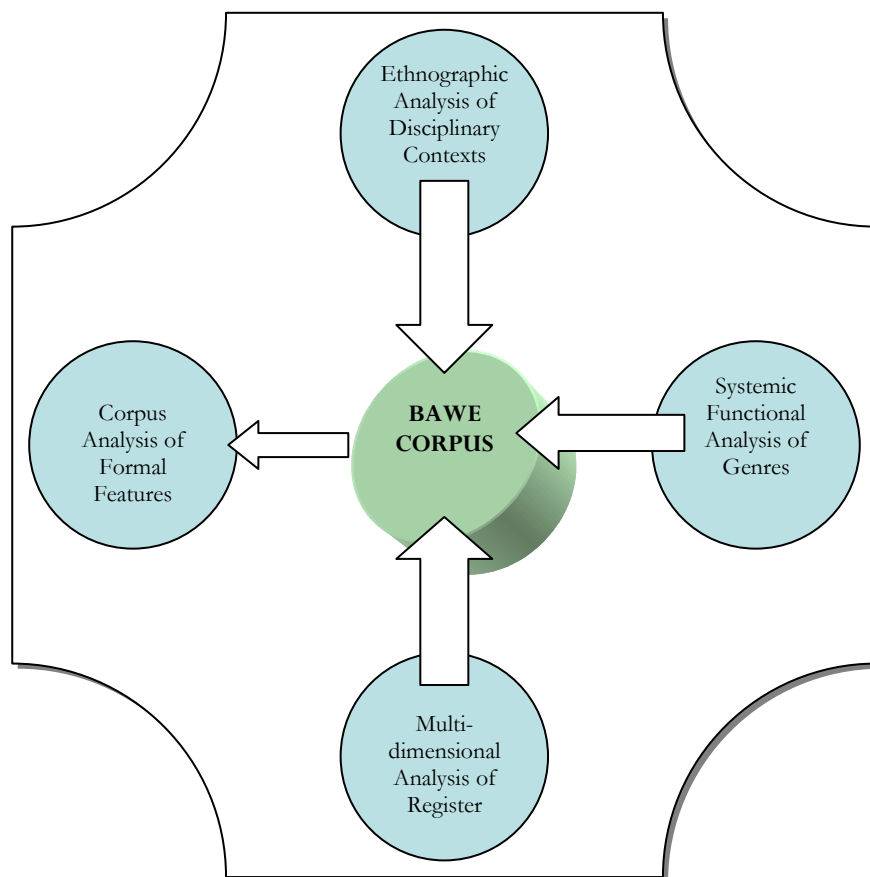
This produced a shift to Stratified Sampling within four broader disciplinary groups and four levels of study corresponding essentially to first, second, and third year undergraduate (bachelor's degree) and first year taught masters.

Table 1: Stratified sampling grid showing level and disciplinary group

	<i>1<sup>st</sup> year</i>	<i>2<sup>nd</sup> year</i>	<i>3<sup>rd</sup> year</i>	<i>4<sup>th</sup> year</i>
<i>Arts &amp; Humanities (AH)</i>				
<i>Life Sciences &amp; Medicine (LS)</i>				
<i>Physical Sciences &amp; Engineering (PS)</i>				
<i>Social Sciences &amp; Education (SS)</i>				

With an eye on automated register analysis of lexico-grammatical features in academic writing in general, others were very much in favour of Random Sampling. But again, the more we considered this, the more we realised it was unworkable. What was our population to be sampled? It would have been beyond our scope (data protection issues aside) to develop a data base of all students likely to produce good assignments, and all assignments they were likely to write (some modules are assessed by numerical or graphical assignments and exams, which are not included in our corpus of 'good' assessed writing). From an ethnographic perspective, we would ideally collect assignments written by students we interviewed, from tutors we interviewed, in departments in which we had spent considerable time and which provided us with explicit course documentation on their expectations of writing in their disciplines. This perspective favoured Cluster Sampling. As illustrated in Figure 1, our desire for a rich account of genres by blending or triangulating theoretical approaches to genre description with different methods of analysis resulted in competing influences on the design of our corpus of British Academic Written English (BAWE).

Figure 1: Competing influences in the BAWE corpus design



This was more than multiple method where different perspectives are ‘triangulated’ on the same data. This was competition among competing theories working with different types of data. Our sampling strategy was inevitably a compromise among these three perspectives and the realities of the collection process. We started with what seemed like a reasonably constrained Purposive Sampling strategy, and planned to interview tutors and students in all our targeted departments. We aimed to collect up to 32 assignments from each of four years of 28 disciplines in four disciplinary groupings, for a total of 3500 assignments and eight million words.

Table 2: Sampling grid of 4 disciplinary groups and 28 disciplines

<i>Arts &amp; Humanities</i>	Applied Linguistics/ Applied English Language Studies; Archaeology; Classics; Comparative American Studies; English; History; Philosophy
<i>Life Sciences</i>	Agriculture; Biochemistry; Food Science and Technology; Health and Social Care; Medical Science; Plant Biosciences; Psychology
<i>Physical Sciences</i>	Architecture; Chemistry; Computer Science; Cybernetics; Engineering; Mathematics; Physics
<i>Social Sciences</i>	Anthropology; Business; Economics; Hospitality, Leisure & Tourism Management; Law; Politics; Sociology

We targeted departments where we had contacts, and avoided those which did minimal writing, aiming to maintain a reasonable spread within each disciplinary group. The issues around disciplinary groupings and disciplines are well documented (e.g. Becher & Trowler 2001), and our decisions here also reflect issues in blending approaches. From an ethnographic perspective we would have asked members of the discourse communities whether they aligned more with social sciences or humanities, for instance. And from this perspective some departments would have been split. We chose here to be influenced by earlier corpus work, in particular the BASE corpus of spoken academic English, which uses the same disciplinary groupings, and the MICASE corpus, its American counterpart, which follows essentially the same groupings but with different labels. This should encourage cross-corpus comparisons.

Some students and some departments were more receptive than others, and this moved us towards more Convenience Sampling. This ensured that researchers had time with individual students to check contextual data for each assignment submitted, thus increasing accuracy. Further details of the multiple collection strategies employed, and the relaxation of the sampling frame towards the end of the project are described in Alsop and Nesi (2008).

Figure 2: Competing influences on sampling strategies

Corpus Analysis of Corpus as Text	→	Random Sampling
Disciplinary Groups and Levels of Study	→	Stratified Sampling
Ethnographically informed investigation of the context	→	Cluster Sampling
Limiting the number of contributions from any given student or assignment	→	Quota Sampling

Gathering of contextual metadata for the file headers in the corpus as a collection of texts	→	Opportunistic /Convenience Sampling
Genre analysis of sets of similar texts in the corpus	→	Purposive/ Judgemental Sampling

Our final holdings reflect the blending of approaches and the resulting compromises. There are sets and strands of the same assignment type amenable to genre analysis in most, but not all disciplines. Multidimensional analysis is possible across disciplinary groupings and years of study, but the number of assignments in some discipline-years is too small for reliable statistical analysis. Our tutor and student interviews match up with some but not all disciplines targeted. It would have been possible to collect our targeted 3000 assignments, but at the expense of balance across categories. Our goals were ambitious, but the first substantial corpus of its kind is now available and others developing similar corpora (e.g. MICUSP) can benefit from our experience. The main value of our corpus now lies in how it is used to further our understanding of academic writing. Before this is illustrated for History and Engineering, two further corpus development issues are discussed.

## 2.2 CONTEXT VS CONTENT IN ASSIGNMENT CLASSIFICATION

In corpora there is usually contextual information in the metadata or file headers which refers to the author and text production context (e.g. age, first language; date of production, discipline). This meant that we should not be using linguistic criteria to identify which discipline a text was from. Questions thus arose about how to define our contextual data in ways that were not circular (Gardner & Hindle 2005).

In British universities the choices open to students following a degree in, say Sociology, are still relatively conscribed and limited, but issues arose over assignments such as sociology for medical students; accounting for cultural policy students; mathematics for economics students. The assignments from the latter module might look like mathematics assignments, but have been written by economics students in a module taught in the economics department. Should we categorise assignments from such modules objectively (and therefore quickly, by one member of the research team) according to the degree course the student was taking, the departmental home of the module, or the tutors teaching it? To analyse it first, and interpret it as belonging to a specific discipline would not only be more subjective, and require more expert classification, it would also, it seemed from contextual metadata perspectives, be distorting the corpus. While a maths for economics students assignment might be relatively easy to classify as mathematical writing, others were more interdisciplinary. For example, psychology-like assignments appear in education, business, medicine and applied linguistics. The interpretive task grew in scope and complexity. We therefore chose to classify assignments according to their home department. This maintained the distinction between the contextual and the linguistic variables used for subsequent



corpus analysis. For most assignments this was not an issue as the student, department, module, and module tutor were all from the same discipline; but the questions again could be traced back to issues of competition among multiple perspectives.

### 2.3 ALL DISCIPLINES ARE NOT EQUAL

The imbalance in number of assignments per discipline in the final corpus can be partially explained by the nature of the disciplines themselves. For example, Medical Science is not taught until level 4 and some departments do not offer masters level modules. Departments vary in size, and some expect a wide range of assignment types, where others are much more homogeneous. Assignments in History are essentially all essays; most students writing History assignments have been writing similar assignments throughout secondary school and have passed A level History. Many during the course of their university degree follow History modules exclusively. Of the 93 undergraduate History assignments in the final corpus, more than two thirds were written by single honours History students; the rest by joint honours students studying for degrees in History and Politics, Sociology, Cultural Studies or French. In contrast, the Engineering assignments are of more than a dozen different types, including projects, proposals, explanations and critiques. Many assignments are lab reports of actual or simulated tests and experiments, while case studies are used in financial and structure of the industry modules; essays in social, legal and ethical aspects of Engineering modules; narrative recounts in modules on starting and running a business, and in project work; and design specifications are popular across a range of design modules. This diversity in module and assignment type contributed to substantially more assignments being collected and the 205 undergraduate Engineering assignments in the final corpus were from students in fifteen different degree programmes including Civil, Electronic, Mechanical and General Engineering.

Over time, our targeted disciplines changed if assignments were not forthcoming. The core–option distinction worked for some disciplines but not others. Modules taken by 2nd and 3rd year students together were classed according to the student, rather than module. All in all, it turned out to be much harder to develop a full and balanced corpus than we initially thought it would be (Alsop & Nesi 2008). Perhaps these comments from Stubbs sum up our current position:

I will ignore here questions of corpus design, since concepts such as 'representative' and 'balanced' corpora make some intuitive sense but cannot be operationalised. A corpus can never be a representative sample of a language since it is not clear what we would want it to represent. Nevertheless, large corpora, which sample widely, reveal major regularities underlying authentic language use (De Beaugrande 2001:113). They allow generalisations to be made, which can be tested on different independent corpora. (Stubbs 2006:17) (emphasis added)

Although the BAWE corpus is not huge by corpus standards, with 2800 texts and over 6.5 million words it is a large stratified corpus of student writing. And while it can never be fully representative, it does lend itself to be used for analyses of academic writing

across disciplines and levels of study. I now illustrate its potential and our blending of multiple approaches through descriptions of History and Engineering.

### 3 ETHNOGRAPHICALLY INFORMED ACCOUNTS OF ASSIGNMENT TYPES

When I began this paper in 2006<sup>2</sup>, the most complete sets of assignments available were in undergraduate years 1-3 History and Engineering, so I decided to focus on these. From Becher and Trowler's (2001) work from a sociological perspective we know that History as a soft-pure discipline is reiterative, holistic, concerned with particulars, and results in understanding or interpretation, whereas Engineering as a hard-applied discipline is purposive, pragmatic, concerned with mastery of the physical environment, and results in products or techniques (2001:36). Soft pure fields place greater importance on creativity of thinking and oral and written expression, while hard pure and hard applied fields place strong emphasis on ability to apply methods and principles (Hativa 1997 in Neumann 2001:138). As disciplines, then, History and Engineering also reflect clear contrasts in disciplinary practice.

In terms of development across levels, Neuman, Parry and Becher point out that

one major difference between the curriculum in hard pure as against soft pure disciplines lies in the cumulative, atomistic nature of the former in contrast to the reiterative, holistic nature of the latter. Hard pure curricula tend to be conceived as linear and hierarchical, building up brick by brick towards contemporary knowledge. ... As against this, knowledge in the soft pure domain [eg History] gives rise to curricula which can be described in Bruner's (1967) term as spiral in their configuration, returning with increasing levels of subtlety and insight into already familiar areas of content. Hence course structures are open and loose ... (2002:407).

Moreover, hard disciplines place greater emphasis on student career preparation (Braxton 1995). This suggests that there will also be differences in progression from years one to three between History and Engineering.

Our investigation of the discourse communities supports these differences. As long-standing members of the same university community, and as applied linguists involved in research on English for academic purposes, Nesi and I had some appreciation of the two disciplines through her earlier tutor interviews for BASE research on lectures and seminars (reproduced in the EASE academic English materials: Kelly, Richards & Nesi 2004; Kelly, Sharpling & Nesi 2006) and through our supervising graduate dissertations on academic writing in these areas. Our investigation of context was also informed by departmental documentation, tutor interviews (Nesi and Gardner 2006), student interviews (Gardner and Powell 2006), and assignment submission forms completed by students in conversation with research assistants. While far from being a full ethnography, our approach was ethnographically informed

---

<sup>2</sup> An earlier version of this paper developed at the University of Warwick and based on corpus holdings at that time was presented at the 19th European Systemic Functional Linguistics Conference & Workshop, 23-25 July 2007, Saarbrücken, Germany

in that we aimed for an emic or insider perspective on academic writing from those involved in its production rather than trying to impose our own understandings. We use the term 'assignment type' and departmental labels to refer to different genres identified by the discourse community.

We thus learned more about how assignments are written:

In year four they have a team project with tasks similar to what they'll get in employment. They work in teams with students from a range of engineering disciplines to tackle various complex engineering problems. They produce a document with costings targeted at a Venture Capital company. (Tutor on Engineering Projects)

about the structure of assignments:

There is a standard structure... They have a layout of headings and a description of the sorts of things that would go under each of those headings, and they map whatever they've done onto that. (Engineering tutor)

and about perceived and anticipated differences between assignment types:

Although a third-year project may be similar in structure to a laboratory report written in the first year, the writer will have to assimilate, evaluate and integrate a wider range of information. (Engineering Tutor)

The purpose of a site investigation report is to inform a client of access, geology, ground conditions, water flows etc. so that the client can decide what sorts of building(s) to plan. Students inspect site, drill boreholes, do tests, then write a report which covers a set list of aspects and presents results. With a factual report, the engineer mustn't do any interpretation; the client does the interpretation. In an interpretive report, the writer also makes recommendations. Professional issues, legal liabilities may be incurred, so there are tight constraints on what may or may not go in. (Engineering Tutor)

The importance of writing in Engineering was supported by Engineering students:

It doesn't matter how good the project is, if it isn't well written, it won't get good marks (Engineering student)

They were also able to explain how they understood the social purpose of the different stages of their texts.

Basically we make a summary first, the reasons are if someone in industry wants to know what we're doing, they haven't got time to read through, they just want to know what's going on. Then the introduction .. Contents .. Then a theory which takes a huge chunk it's explaining everything... (Engineering student)

This type of information is very useful for genre analysis.

Drawing principally on interviews and departmental documentation, we established that Engineering students write the following types of assignment:

- Laboratory reports
- Project reports
- Reflective journals
- Posters (e.g. for transport museum)
- Site investigation reports (both factual and interpretative)
- Funding proposals
- Business plans
- Essays

Our investigation of writing in Engineering showed that it is leading in many of the cross-disciplinary trends we observed, namely in reflective writing, in writing for non-specialists (e.g. newspapers, museum visitors), in writing for professional purposes (e.g. for industrial funders), and more ICT related writing (e.g. blogs and webpages) (Nesi & Gardner 2006).

In stark contrast, writing in History focuses almost exclusively on the essay. Tutors distinguished formative or class essays of 2000-2500 words written in response to a question and described as ‘a rhetorical exercise to persuade the reader to accept the argument’ from summative essays of 4000-5000 words which are similar, but whose grades count towards the degree classification. Longer essays of 8000 words, it was explained, ‘require additional organisational skills’.

History students explained how they craft their essays:

I try to use different opinions. I say someone’s opinion, then counter it with someone else’s. I weave my own perception in but I’d never say “this is what I think” directly. I use some arguing and counter but I always go back to my introduction stance. (History student)

... I put across my argument always and a. you have to consider the popular argument at the time. You don’t want to go against the flow completely as you don’t have the skill to do that. But b. I consider what the professor will think. (History student)

So History assignments were described in very different ways than in Engineering. There was some sense that more variety was anticipated; for instance, one student had submitted his formative ‘essay’ as a play; and there had been some moves recently to diversify, but the essay remains the mainstay. We established that History assignments were of five main types:

- Formative Essays (2000-2500 words)
- Summative Essays (4000-5000 words)
- Literature Review (2500 words, recently introduced in one core module)
- Book Review (in one optional module)
- Summative Essay (8000 words; optional)

Thus while there was evidence of different types of assignment, the staple in History is the essay.

Our ethnographically-informed account of the discourse community perspectives not only presents one account of genres of assessed student writing, but also provides insights which informed and in some cases contradicted accounts from other perspectives.

#### 4 MULTIDIMENSIONAL ANALYSIS OF REGISTER IN THE CORPUS AS TEXT

The most quantitative aspect of the project is the multidimensional analysis (MDA) developed by Biber (1988). As Xiao and McEnery point out

MDA is undoubtedly a powerful tool in genre analysis. But associated with this power is complexity. The approach is very demanding both computationally and statistically in that it requires expertise not only in extracting a large number of linguistic features from corpora but also in undertaking sophisticated statistical analysis (2005:63).

We therefore sent the BAWE corpus to Biber and his team in Arizona where it was analysed both according to existing dimensions, and independently to generate BAWE specific dimensions.

Basically, MDA as explained in Conrad and Biber (2001:Chapter 2) involves tagging lexico-grammatical features identified as potentially important. The original study involved tagging 67 features such as three tense and aspect markers ('past tense', 'perfect aspect' and 'present tense'); nominalizations ending in *-tion*, *-ment*, *-ness*, *-ity*; public verbs (*complain*, *explain*, *promise*), private verbs (*believe*, *think*, *know*), and suasive verbs (*command*, *propose*, *recommend*). Following tagging, frequencies are calculated, and normalised per 1,000 words. Then factor analysis groups salient factors which are interpreted so that dimensions of variance are identified which allow comparison among registers. The assumption is that "particular sets of linguistic features co-occur frequently because they serve related communicative functions." (Conrad and Biber 2001:24) Dimension scores (or factor scores) can be computed for each text by summing the frequencies of the features having salient loadings on that dimension.

The BAWE corpus was analysed first using the dimensions identified in the original studies, and secondly to generate BAWE-specific factors (see *Appendix*) that should prove particularly relevant to academic writing. I present here some findings from the first analysis, using the 1988 Dimensions originally interpreted for a range of spoken and written genres. For this analysis, 2800 texts in the BAWE corpus were analysed by level, disciplinary group, discipline and genre family; Electronic Engineering (EE) and Engineering (Eng) were treated separately.

Dimension 1 differentiates Biber's face-to-face conversations (+35) and academic prose (-15) through features such as nouns, attributive adjectives, pronoun *it*, and word length. Interestingly the results show significant differences between the groups of texts at each level in the BAWE corpus. This might be attributed to increasing word length from Year 1 to 4. There is no significant difference between results for the Disciplinary Groups of PS and AH, although there is a significant difference between Electronic

Engineering and History, with History falling at the lowest end of this Dimension with a mean of -19 alongside Comparative American Studies, Medicine and Biological Science.

Table 3: Dimension 1: Involved vs Informational Production

Level	N	Mean	Disciplinary Group	N	Mean	Discipline	N	Mean
Year 1	814	-13	AH	704	-13	Electronic Engineering	28	-12
Year 2	779	-14	PS	608	-13	Engineering	246	-14
Year 3	559	-15	SS	790	-15	History	96	-19
Year 4	648	-17	LS	698	-16			
Total	2800		Total	2800				

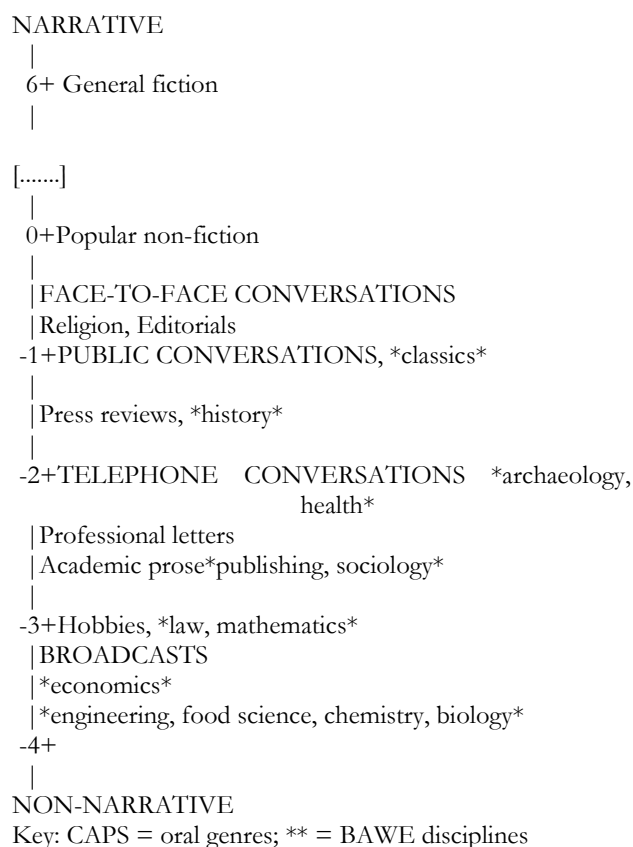
Dimension 2 is the Narrative Dimension, calculated through features such as past tense verbs, third person pronouns, and synthetic negation (e.g. no evidence was found) which differentiates Romance Fiction at more than +7, from Academic Prose at less than -2. In the BAWE data, there are significant differences between AH and PS in Dimension 2, and between History and both Engineering groups.

Table 4: Dimension 2: Narrative vs Non-Narrative Dimension

Level	N	Mean	Disciplinary Groups	N	Mean	Disciplines	N	Mean
Year 1	814	-2.7	AH	704	-2.1	History	28	-1.6
Year 2	779	-2.8	PS & E	608	-3.0			
Year 3	559	-3.0	SS	790	-3.0	Electronic Engineering	96	-3.7
Year 4	648	-3.2	LS & M	698	-3.7	Engineering	246	-3.8
Total	2800		Total	2800				

As Conrad (2001:103) in a MDA study of History textbooks and research articles points out: “All the history texts focus on human events, and, by definition, history is concerned with the past. It is not surprising then that history texts would make use of past tense and third-person pronouns to refer to people and the past.” When we plot this on Biber’s graph, and add further BAWE disciplines, we see that history essays are less ‘narrative’ than general fiction, however.

Figure 3: History and Engineering Student Writing on the Narrative Dimension



Dimension 3 is the Situation-dependent vs Elaborated Reference dimension that differentiates Official Documents (+7) from Radio Broadcasts (-9) through positive features such as wh-relative clauses and nominalisations and negative features such as time and place adverbials. This dimension finds no significant difference between Electronic Engineering (+3.9), Engineering (+4.5) and History (+5.9). All are towards the elaborated end of this dimension.

Dimension 4 and Dimension 5 find significant differences between Genres that are frequent in History and Engineering. Dimension 4 is the Overt Expression of Persuasion/ Argumentation Dimension which differentiates Editorials (+3) from Broadcasts (-4). It groups Proposals (1.3), Design Specifications (0.7) and Case Studies (-0.5) as significantly different from Critiques (-1.6), Essays (-1.8), Research Reports (-2.4) and Methodological Recounts (-2.5) through positive features such as infinitives, modals, and suasive verbs.





through this bespoke BAWE multidimensional analysis are the groupings of disciplines as similar along specific dimensions.

Table 5: Disciplines grouped with History in the Biber Pilot<sup>3</sup> Corpus

Factor 1	History, Psychology, Anthropology, Classics, English
Factor 5	History, Psychology, Anthropology, Law

The potential of this type of empirical research to reveal similarities and patterns hitherto unnoticed is great. For instance, this pilot data suggests similarities between writing in History, Psychology and Anthropology which are worth investigating further. To make sense of them, however, we have to use more interpretive approaches. This process is similar to ‘shunting’ which Miller attributes to Halliday:

This notion of shunting, or spread, in analysis is put forward by Matthiessen [2006] as an activity legitimated by Systemic Functional Linguistics (SFL) theory, one involving the analyst’s move in different directions within its multidimensional<sup>4</sup> model, between and across instantiations but also levels of stratification. Halliday (1961, in 2002:45) is even more peremptory: ‘[...] “shunting” is a descriptive method that is imposed on description by theory’. (Miller 2006:267)

For Halliday 1961 (in 2002, pp. 56 and 69) shunting is ‘moving up and down the rank scales’, or in our case moving from lexico-grammar to register and genre. Keeping within the theory, SFL scholars might lay themselves open to the corpus linguist’s critique of not noticing certain patterns of interest; Biber’s multidimensional analysis allows us to start with large multifactorial analyses of the lexico-grammar, with potential to reveal unexpected regularities. To make sense of the patterns revealed, however, we need a complex theory of language, such as SFL.

## 5 CONVENTIONAL CORPUS LINGUISTIC ANALYSIS

The more widespread methods used to analyse text in corpora stem from the ease with which computers can count data strings. They include counts of sentence length, paragraph length, and text length, all of which are available for the BAWE corpus.

<sup>3</sup> This Biber Pilot corpus is different from the BAWE Pilot Corpus developed prior to 2004.

<sup>4</sup> This use of *multidimensional* does not refer to Biber’s multidimensional analysis

Table 6: Automated Counts of Corpus Identifiable Features

DISCIPLINE	History				Engineering			
LEVEL	1	2	3	1-3	1	2	3	1-3
number of texts	30	32	31	93	63	75	67	205
<b>Means by Assignment:</b>								
length in words	2371	2694	2886	2654	1711	2023	3195	2310
number of sentences	86	86	92	88	87	107	159	118
number of paragraphs	15	14	16	15	41	52	73	56
sentence length in words	29	34	33	33	21	20	20	20
paragraph length in sentences	6.3	6.4	6.1	6.4	2.5	2.6	2.5	2.6
number of tables					1.6	1.4	2.4	1.8
number of figures	0.1	0.1		0.1	3.3	3.7	8.7	5.2
number of block quotes	0.4	0.6	0.7	0.6	0.4	0.4	0.4	0.4
number of formulae			0.03	0.01	9.3	16.0	20.0	15.3
number of lists					1.8	1.9	1.7	1.8
number of listlikes					4.8	5.6	8.6	6.7

As tables, figures, block quotes, formulae and lists are tagged in the BAWE corpus (Ebeling & Heuboeck 2007), descriptive statistics are readily available for these features. We see from such information where there is development from year 1 to year 3, as in text length, and where there is not, as in sentence and paragraph length. More striking are the differences between writing in History and Engineering in terms of tables, figures, formulae and lists. Viewed in the context of the full corpus, we see that history and engineering are at different extremes in overall means on many of these measures. For example, the mean length in words for year one of the whole corpus is 1788 (between 2371 and 1711); the mean sentence length in words is 25 (between 29 and 21) and the mean paragraph length in sentences is 4.8 (between 6.3 and 2.5).

Such descriptive statistics provide information that otherwise would not be readily accessible. When we compare the statistics with the range of text types in Engineering, it makes sense that there are more features such as figures in comparison with the essays in History. Programmes such as Wordsmith Tools (Scott) or SketchEngine (Kilgariff) are readily available to examine type-token ratios, word frequencies, collocations, ngrams or lexical bundles, and key words. The motivation for such computations comes from the ease with which computers can process and reveal patterns in large amounts of text, but also from the search for stronger empirical support for patterns observed in individual texts.

Mathiessen (2006:110) describes this as a ‘two-pronged approach’ which typically starts with manual analysis of a ‘small’ sample of text, and then extends to cover a larger sample of text for certain aspects of the analysis that can be stated in terms of patterns of orthographic words. He points out that

The general methodological point is that the view ‘from below’, based on graphological patterns, only allows us to see a relatively small part of what can be analysed manually ‘from above’. This is a key reason for adopting the two-pronged approach involving manual analysis of small samples and automated analysis of large corpora. (Matthiessen 2006: 112-3) (my emphasis)

An example of this for History in our project is the manual analysis of Field in the initial sentence subject of assignments in the corpus which gives rise to various hypotheses about how different disciplines demonstrate progression from levels one to three (Gardner, in press). Such manual analysis can be further investigated through automated searches, as the Wordsmith Tools data in Table 7 shows:

Table 7: Evidence for Progression in History from Phenomena to Metaphenomena

	Rank	Raw Freq	Freq/1000wds	Texts	%Texts
‘History’ Yr 1	74	75	1.3	15	75
‘History’ Yr 2	52	151	1.8	16	73
‘History’ Yr 3	24	304	4.1	14	87
‘Historians’ Yr 1	172	34 (+5) = 39	0.6 (+0.0) = 0.6	9 (2)	45 (10)
‘Historians’ Yr 2	204	45(+11)= 56	0.5 (+0.1) = 0.6	13 (7)	59 (32)
‘Historians’ Yr 3	139	52(+33)= 85	0.7 (+0.4) = 1.1	8 (6)	50 (37)
(+ ‘Historian’)					

This table shows an increase in the use of the terms *history* and *historian(s)* from first to third year of study in the 2006 corpus of 60 History texts. The evidence for increase in use of *history* is consistent across rank in word frequency (74<sup>th</sup> to 24<sup>th</sup>), mean frequency per 1000 words (1.3 to 4.1) and occurrence across texts (from 75% to 87% of texts in the year group). Similar trends are found in an examination of *historian* and *historians*, which occur less frequently. This provides empirical evidence in support of the claim that students write less about the events in history (phenomena) and increasingly about the discipline itself (metaphenomena) as they progress towards graduation.

Where Mattiessen takes the view that corpus analysis can only reveal a very small part of the picture, Stubbs takes a rather different point of view and argues that many paradigmatic and syntagmatic relations are only visible through concordancing.

In an individual text, neither repeated syntagmatic relations, nor any paradigmatic relations at all, are observable, but a concordance makes visible, at the same time, what frequently co-occurs syntagmatically, and how much constraint there is on paradigmatic choice. These repetitions objectively exist. (Stubbs 2006:18)

Potentially more interesting than cases where approaches offer mutual support to hypotheses, are those cases where contradictions emerge. It is here that the researcher has to probe for fuller understandings and clarity. One example of contradictions between multiple accounts lies in assignment word length.

On the surface it might appear that assignment length should be a feature most amenable to empirical computational analysis, and yet our multiple perspectives extend to both units of analysis, ‘word’ and ‘assignment’. When we compare the word length of History essays in Table 6 with the contextual information provided by the disciplinary community in section 2, it appears that the means in Table 6 may be hiding the range of assignments in each group. This turns out to be the case, and assignments in Year 3 range from 1230 to 9090. This suggests a very different picture of increase in assignment length than the modest increase in means in Table 8.

Another contradiction arises in determining how much of the assignment should be included in the word count. Did we really want to include words in the appendices, for instance? In assignments with references and appendices, this can make a difference:

Table 8: Word Length Example

MS Word word count	MS Word text file count	Corpus as Mass Word Count	Student attributed Word Count	Disciplinary Context Word Count
5341 words	5234 words	4867 words	4861 words	5000 words

In Table 8 there are five different word lengths for this one essay (0003k). The Word file count of 5341 words was prior to processing and included the student’s name, tutor’s name, university name and any other details that might identify the student which have been removed for data protection reasons in preparing the file for the corpus, as indicated by the text file length of 5234. The corpus word count of 4867 is based on running text, which is the whole file minus tables, figures, and front and back matter such as the title page, references and appendices. As these often contain rather different writing from the running text, these were blocked by taggers from word counts. This assignment had several hundred words of front and back matter. It is interesting that in this instance the student also presumably counted the word length of 4861 as excluding the front and back matter. The final word count (5000 words) is how the assignment length is described by the disciplinary community in handbooks and interviews. Such contradictions between perspectives usefully force us to clarify our definitions and scope.

From the perspective of the disciplinary context, an assignment is something that is assigned, submitted and potentially awarded a grade. As we discovered, this might correspond to a collection of essays, or a collection of reports, sometimes with a common introduction and references. For genre analysis, there was interest in analysing

such collections as compound assignments, and making the individual text parts amenable to further genre analysis with other essays or reports. The BAWE corpus thus consists of 2761 assignments, which correspond to 2800 texts in the Biber analysis when compound assignments are unpacked. From the corpus perspective, it was easier to keep the unit of analysis as the whole assignment; this was, after all, the main focus of our investigation.

## 6 GENRE ANALYSIS OF THE CORPUS TEXTS

Genre analysis is not only central to our investigation, but also a daunting task if viewed as detailed analysis of several thousand texts. If we are to be able to interpret computational data such as concordance lines or frequency counts in meaningful ways, we want to be able to locate these features in different genres and at different stages of texts. To date, our mark-up and search engine capacities enable us to refine searches of the entire corpus by genre family, discipline, disciplinary group, and level. We can thus compare, for example, feature of History year one with History year two; or essays in Arts and Humanities with essays in Social Sciences. The corpus has been tagged so that searches can identify features at different section levels, in different paragraphs (e.g. 2/9, the second of nine paragraphs) and at different places in the paragraph (e.g. 1/3, in the first of three sentences). Such ready production of empirical data can assist in genre analysis. For example, extracting assignment titles and section headings can inform descriptions of assignment macrostructures (Gardner & Holmes, forthcoming), which provide excellent initial indicators of genres and their generic stages. In contrast, the analysis of the corpus texts into thirteen genre families or groups of similar genres (Gardner & Nesi 2008) was completed manually. The identification of genre families was planned to allow us ready comparisons across disciplines. For example, in our genre family of Proposals, we have research proposals, book proposals, marketing plans, legislation reforms, design proposals and catering plans. These assignments share the purpose of demonstrating the students' ability to make a case for future action, with three main generic stages labelled purpose, detailed plan and persuasive arguments. They differ from Design Specifications whose purpose is to design a product with generic stages of design brief, design considerations, optional testing and a design plan. There are design specifications in the corpus for applications, products, websites, systems labels, buildings and games.

Interviewees commonly emphasize the novel and the unusual, perhaps in an attempt to engage or impress the interviewer. Our analysis into genre families puts this in a numerical context.

Table 7: Genre Families in History and Engineering

DISCIPLINE	History				Engineering			
YEAR	1	2	3	1-3	1	2	3	1-3
percent of assignments with abstract	0	0	0	0	44	47	48	139
<b>Genre Families:</b>								
Case Study					10	2	7	19
Critique			1	1	4	5	12	21
Design Specification					6	11	17	34
Empathy Writing					0	1	0	1
Essay	30	32	30	92	1	3	4	8
Exercise					1	2	8	11
Explanation					8	4	3	15
Literature Survey					1	0	0	1
Methodology Recount					30	35	10	75
Narrative Recount					1	6	1	8
Problem Question					0	2	2	4
Proposal					1	3	0	4
Research Report					0	1	3	4
TOTAL	30	32	31	93	63	75	67	205

This confirms that essays are the mainstay of History, where Engineering students write many different types of text, with a large share of them being laboratory reports of some kind (methodological recounts). It is also clear from this comparison that the corpus does not include examples of all assignment types. This may be attributed to the nature of the assignments excluded (e.g. posters with minimal writing) and the emergent nature of some assignment types mentioned (e.g. literature reviews in History).

## 6.1 PROGRESSION IN HISTORY

While the best clues to Engineering genres lie in the Macrostructure and headings, the titles of History assignments are often good initial indicators of genre.

### TITLES EXPECTING DISCUSSION:

'The Impact of the Black Death has been Greatly Exaggerated.' Discuss.  
Was migration into early modern London driven more by 'push' or 'pull' factors?

### TITLES EXPECTING A THESIS:

Why did the East Asian economy grow more slowly than the European economy by 1800?  
How should we explain the gap between theory and practice in the social position of women in early modern England?

TITLES EXPECTING AN EXPLANATION:

What factors account for the failure of the Church Missionary Society's Wellington Valley Mission?

What part did the factors outlined in Weber's Protestant Ethic play in explaining the divergent economic development of Western and Eastern civilisations?

It was of course possible for students to produce a thesis at the beginning of their response to a 'discussion' type question, but usually this did not happen. Following work by Coffin (e.g. 2004) and others on secondary school history, it was surprising to me how readily university assignments fitted the genres identified. Here is an overview of five of Coffin's history genres and their generic or schematic structure, with optional stages in brackets.

ANALYTICAL EXPOSITION (AE): (Background)^ Thesis^ Arguments^ Thesis Reinforcement

ANALYTICAL DISCUSSION (AD): (Background)^ Issue^ Arguments^ Position

ANALYTICAL CHALLENGE (AC): (Background)^ Position Challenged^ Arguments^ Anti-Thesis

FACTORIAL EXPLANATION (FE): Outcome^ Factors^ Reinforcement of Factors

CONSEQUENTIAL EXPLANATION (CE): Input^ Consequences^ Reinforcement of Consequences

These stages are now illustrated from History assignments, with bold added to highlight key linguistic features.

BACKGROUND in history is typically narrative written in the simple past, with temporal markers, such as Jean-Antoinette Poisson was born in .... on 29 December 1721, .... re-christened the Marquise de Pompadour, she became the *official royal mistress to Louis XV in 1745*, .... (AD). It usually appears at the beginning of history essays; essays in English also use this, but less frequently. They are equally likely to start with a quote or reflection.

THESIS states the main argument the student wishes to make and often includes an overview of how the essay is structured to support this argument. It usually appears towards the end of the first paragraph or first section of the essay. For example, I **will attempt to show why I agree with** scholars such as Craig Calhoun **that it was** primarily artisans that resisted the changes brought by industrialization and who seemed to have more of a sense of unity **in comparison to** the factory workers. (AE)

ISSUE introduces the focus of discussion, and like the thesis usually often occurs towards the end of the first paragraph or section, with an indication of the order of the arguments. ***This essay will look at continuity and departure through these two periods thematically, and attempt to form a conclusion as to whether the 'Age of Braudel' conserved or contradicted original Annales historiography.*** (AD)

ARGUMENTS in History tend to be realised in one or more paragraphs. They may be grouped under a MacroTheme (Martin and Rose 2003) that introduces a series of arguments, such as *one first needs to find evidence supporting the idea that ...* (AD). The arguments may then be marked in support or contrast to the previous argument:

*The traditions of the artisan, **too, meant that** they already ...  
**Furthermore,** the Revolution ....  
**Another reason why** artisans were more likely to ...  
**However despite the above reasons as evidence for** 'radical artisan' discourse, it  
 is **not** so simple to ....  
**There is also criticism of** ...  
**As well as seeing why** artisans ... **it is important to look at why** ... (AE)*

Argument paragraphs often begin with a statement of topic (hyper-Theme), proceed to add detail and comment, and conclude with the contribution of the paragraph to the larger argument of the assignment (hyper-New) (Martin and Rose 2003). Here are some examples of Hyper-Themes and Hyper-News:

***A prime characteristic** of modern environmentalism **was** an endeavour to conserve exhaustible sources of energy by conducting searches for safe, renewable sources. .... **This clearly represents how** modern environmentalism adapted qualities from the path set down by Romantic Movement as both encouraged the return of man to nature. (AD)  
 In attempting *histoire totale* the founding fathers of the *Annales* were rationalists: 'whatever can be rationally pursued, measured, calculated, quantified, is so pursued'. Block favoured ...  
**Thus** in terms of his scholarship and the leadership of the institutions of the *Annales*, Braudel conserved the quantitative tradition. (AD)  
 Neither **is** it wholly acceptable **to maintain that** hunger develops when population growth overtakes the capacity to provide subsistence, as in India .... **,thus** poor people should not be held accountable for their abject condition of hunger.  
 When Germany attacked Moscow in 1941, it appeared that .... **This illustrates that** left to their own devices the Russian people could not defend the capital.  
**There are** currently vast swathes of unoccupied land that have not been invested in for purposes of cultivation; only ... **Therefore,** the stubborn idleness of peasants **can be seen as a reason for** why many people today go hungry ... (FE)*

The Hyper-Theme is often the first sentence in a paragraph, but it may also be preceded by a linking move: *In order to affirm Madame de Pompadour's popularity at Versailles, one needs more evidence of affection than merely shown by the king; ... The obvious place to look **next is** to the royal family... **It would seem therefore that,** if Madame de Pompadour was truly popular at Versailles, this did not derive from favour among the royal family, aside from her lover. (AD)*

POSITION The position stage sums up the conclusions of the arguments as developed in the hyper news: *Braudel clearly continued .... Yet it is in ... that the greatest divergence can be found ... **Thus is may be possible to talk of** two Braudels: (AD)*

Similarly, the REINFORCEMENT OF THESIS stage sums up the arguments made and restates the elaborated thesis. *Therefore, as women were neither utterly subservient nor equals, **the central task of this essay has been** to outline how servile they were. Variables ... Generally however ....**Nonetheless,** the extent to which women had authority in early modern marriages was ultimately controlled by their husbands. (AE)*

***Thus in conclusion I would argue** that the novel **is indeed** about degeneration as .... (AE)*

Factorial Explanations aim to 'explain the reasons or factors that contribute to a particular outcome' (Coffin 2004:270). The initial OUTCOME stage introduces the outcome and previews the reasons or factors to be explained, as these extracts suggest:



*This essay will account for and analyse these separate reasons and come to the conclusion that the vast difference in belief systems was undoubtedly the primary cause for the failure of the mission. (FE). The reason that this caused such radicalism in the Russian labour movement .... can be attributed to many factors. (FE)*

The FACTORS, like the arguments are linked together to build towards a full explanation of the outcome using language such as: *First, ... Another important factor which contributed .. is... ; However, perhaps the most important reason for ... is probably the most obvious reason; that ....; Ideologically, there then seems to be a number of distinct factors determining economic development (FE) ... ; Biological and environmental factors are also not rejected by Weber .... ; We can challenge, however, the extent to which .. these factors outlined within the Protestant Ethic do adequately explain the divergence .... (FE); One of the principal factors in explaining the failure of the mission was the fact that the Aborigines ... used missions for purposes other than religion (FE)*

These factors are then brought together in the final REINFORCEMENT OF FACTORS as suggested in these extracts: .... , *a whole number of factors then playing a part in explaining .... ; factors which still to this day are debated and analysed by a variety of contemporary commentators. (FE)* *Therefore, in conclusion, there were a huge variety of factors which explain the failure of .... Some of these were specific to ... However, one factor stands above all others in explaining the failure .... : the vast differences in ... meant that the ... was doomed to failure from the start. (FE)*

In a Challenge, the essays usually start with an account of the POSITION CHALLENGED (AC), and may or may not include an antithesis in the same stage: *An attempt will be made, therefore, not only to dispute much of Macpherson's argument, but to propose a different hypothesis of how the Levellers interpreted franchise reform, based around notions of compromise and heterogeneity. (CH)* The ANTI-THESIS is either reinforced or stated in this final stage, following the arguments: *The aim of this essay has been not only to highlight many of the problems of MacPherson's thesis, but also to propose an alternative interpretation of Leveller franchise reform proposals. ... Not only does this highlight the main problem with MacPherson's thesis, but illustrates one of two factors that this essay has stressed as vital in understanding .... What also needs to be stressed is ..... They proposed an incredibly radical and 'modern' set of ideals that ...*

In addition to challenging published sources, students occasionally challenged the assumptions of a question set by a tutor. This would be part of the development of the thesis or position stage, rather than the main organising principle of the essay.

In a 2006 study of progression, I analysed 60 assignments from 21 different modules on 59 different topics written by 17 different young (18-22) native English speaking students following a BA course in History. This suggests that developing arguments is the mainstay of university history writing, as Arguments is an obligatory stage in Exposition, Discussion and Challenge genres.

Table 8 Progression in History Genres

Yr		Analytic Exposition	Analytic Discussion	Factorial Explanation	Consequential Explanation	Analytic Challenge	Book Review
1	22 essays	7	10	2	2	1	
2	22 essays	9	9	1	1	2	
3	15 essays 1 critique	7	3	3	1	1	1
	60	23	22	6	4	4	1

Where Coffin had found a progression from narrative to analytical genres in her study of secondary history texts, I had looked for a decrease in Factorials and an increase in Challenges as students progressed through university. However no progression in genre is evident from this sample.

Following Coffin's use of Wordsmith Tools, I also looked for evidence of progression from narrative to argumentative language through Years 1-3, and found no clear evidence. As we saw from the MDA, the History register is narrative, but the descriptions of historical events form background information or are construed in support of arguments. Narration is therefore not the main purpose of university history essays; it is constructing arguments.

Drawing on the interview evidence I looked for a 'literature review' stage in the long essays. There certainly was more reference to more literature in these essays, and further investigation showed that there was also a qualitative increase from Year 1 to 3 in terms of the nature of sources: from single authored books in Year 1, to include edited books in Year 2, and occasional journal articles in Year 3. There was, however, limited evidence of a clear literature review stage, and not enough for me to propose a new genre. Further work on Year 4 history essays might prove worthwhile in this respect.

Further study of progression in History should also focus on the central arguments. We have evidence that the nature of sources moves from single accounts to multiple authors and would expect the writing to be increasingly dialogic. This supports the evidence from this ISS analysis of field of a move from a focus on phenomena to metaphenomena.

## 6.2 ENGINEERING GENRES

Although there has been work on Engineering Genres, there was not a similar set of pedagogic genres ready to fit the student texts. We knew from the discourse community the main assignment types, and identified the following key genres in our corpus of texts:

Table 8: Main Genres and Genre Families in Engineering

Genre Families	Genres	Purpose
Methodology	1. Lab report	To test...To discover...
Recount	2. Design report	To design and test
Exercise	13. Calculations 14. Short Answer	To practise or demonstrate competence in discrete skills
Explanation	3. Design concept overview 4. Industry overview 5. System overview	To explain a concept, entity or system
Critique	6. Evaluations of products, techniques, performance, systems, tools, and buildings	To demonstrate understandings of and the ability to evaluate and/or assess the significance of the object of study
Case Study	7. Company Report 8. Accident Report	To assess a company and recommend an investment strategy To review an accident and recommend prevention strategies
Design Specification	9. Design Plan	To develop a design to meet an engineering brief (eg a specific tool)
Proposal	10. Design Proposal	To persuade reader to contract the design team
Essay	11. Exposition, Discussion	To demonstrate the ability to develop an argument
Narrative Recount	12. Reflection on Team Work	To document how team members worked as a group and reflect on their contributions
Research Report	15. Project	To demonstrate ability to carry out a complete piece of research and relate its significance to the field

Lab Reports include testing properties of different materials, comparisons of manual and computerised modelling; simulated and practical experiments. Although there is a Theory section, this is very short and is not the focus of the assignment. The general structure of lab reports is familiar to students, as is the register: *In conclusion it was found that the friction factor of a pipe depends largely on the wall shear stress inherent in the flow (and 6 other bullet points)*. The typical Lab Report headings are Summary^ Introduction^ Theory^ Apparatus and Methods^ Observations and Results^ Analysis (Discussion)^ (Recommendations)^ Conclusions^ (References)^ (Appendices). The aim might be “*to model data describing both the longitudinal and lateral force characteristics of a road tyre and then to evaluate the linear and non linear models used to analyse the data.*” As here, there is generally some evaluation of method and suggestions for how the experiment could be improved. Design Reports have the same generic stages as Lab Reports, but the Theory section is more developed. Thus they aim to design and test. These are both typical of many Engineering genres in their inclusion of tables and figures, (cf Table 6) as this extract from the Apparatus section of one Lab Report suggests:

*Figure 6 shows the flow diagram of the system. Table 1 shows where the pressure is measured relative to the face end of the pipe. Figure 7 is a detailed view of the pipe tappings, intake and outlet. The apparatus consists of a horizontal pipe with tappings along the length of it. One end ....*

Exercises were often calculations which provided short practice in engineering calculations. These genres develop core engineering methods.

Explanations in Engineering describe entities from theoretical concepts, such as road vehicle drag, to complex systems, such as the structure of the UK car industry. Their aim is to demonstrate understandings of the components, workings and roles of the object being explained. Critiques include Evaluations of products, techniques, performance, systems, tools, and buildings. They include a description of whatever is being evaluated, and evaluations. This may be comparative (e.g. manual vs computerised); may be individual (e.g. strengths and weaknesses); and may involve testing parts or functions (e.g. the performance of a domestic solar collector water heating system). Case Studies are multifaceted reviews of complex organisations or events with recommendations for future action. Two examples from Engineering are company reports and accident reports. The former includes investment recommendations and the latter accident prevention recommendations. These genres developed understanding of the role of engineering systems in real or simulated contexts.

As emerged from the interviews with tutors and students much assessed writing in Engineering aims to produce professional engineers. Design Plans specify a design for a tool, for instance, in response to a specific Engineering design brief such as to design a torque sensor. Problem Questions ask students for solutions to realistic professional problems; Design Proposals are written with costings as if to persuade a Venture Company to invest: “...For an initial expense, the investment into CAD and FEA is of great benefit to enterprises...” They are set in a context, such as designing smart homes for the elderly (i.e. with various sensors), and typically include a Summary^ Introduction^ Design^ Costs^ Conclusions. This requires students to follow professional structures in their writing, and to become aware of constraints imposed on professional engineers, although as Freedman *et al.* (1998) and others have shown, these are still university assignments, subject to grading by tutors.

Essays are used in the ethics modules in response to questions such as ‘The primary goal of a course on engineering ethics ought to be to have students master the standards of professional conduct specified in the major engineering codes of ethics.’ The Reflective Pieces are narrative accounts of group project work, in which individual students analyse the roles they took, how the group functioned through the project, and then evaluate this performance. They tend to be more reflective than the Project self evaluations which are similar to limitations in published research:

*the engineering data obtained ... was very limited .... The work is important as it showed .... The work should be continued to determine .... The work was a success in that ... The extent of the experimental work completed is not as great as originally expected ... The author found obtaining information on ... very difficult ... There was no scope in the available time to consider alternative .... Whilst efforts were made to minimise variations in ... these could not be eliminated completely during the project phase.*

As this suggests, the Project itself is a substantial piece of work (around 10,000 -15,000 words but with formulae, figures, graphs and tables, making 50-60 pages) that pulls together most of the skills learned in other modules. It includes an Abstract or Summary, Self Evaluation, Table of contents, Introduction, Literature Review, Theory, Methods, Observation and Results, Analysis and Discussion (with optional costing), Conclusion, References, and Appendices. The stages of the final research report have been practised in isolation in earlier genres, and it is in this research report that they come together.

### 6.3 PROGRESSION IN ENGINEERING GENRES

Table 9 presents the distribution of genre families for levels 1 to 3 in the corpus. It should be remembered that this does not exactly reflect the distribution of assignments written, but nevertheless the relative decline of MR and Explanation together with the increase in Design Specifications and Critiques suggest a clear move away from the factual towards the more creative and evaluative. Although the numbers in some categories are small, the genre analysis is supported either through recognition of component parts in earlier genres, as in the case of research reports; or in recognition of similar genres from other disciplines, as in the case of problem questions. This again demonstrates the value of a large corpus of student writing for developing a trustworthy account of student writing across the disciplines.

Table 9: Engineering families by level

	1	%	2	%	3	%	ALL	%
Methodology Recount	30	<b>48.4</b>	35	<b>46.7</b>	11	<b>16.2</b>	76	<b>37.1</b>
Explanation	7	<b>11.3</b>	4	<b>5.3</b>	3	<b>4.4</b>	14	<b>6.8</b>
Design Specification	6	<b>9.7</b>	11	<b>14.7</b>	18	<b>26.5</b>	35	<b>17.1</b>
Critique	4	<b>6.5</b>	5	<b>6.7</b>	11	<b>16.2</b>	20	<b>9.8</b>
Research Report	0	<b>0</b>	1	<b>1.3</b>	3	<b>4.4</b>	4	<b>2</b>
Problem Question	0	<b>0</b>	2	<b>2.7</b>	2	<b>2.9</b>	4	<b>2</b>
Case Study	10	<b>16.1</b>	2	<b>2.7</b>	7	<b>10.3</b>	19	<b>9.3</b>
Narrative Recount	1	<b>1.6</b>	6	<b>8</b>	1	<b>1.5</b>	8	<b>3.9</b>
Essay	1	<b>1.6</b>	3	<b>4</b>	4	<b>5.9</b>	8	<b>3.9</b>
Exercise	1	<b>1.6</b>	2	<b>2.7</b>	8	<b>11.8</b>	11	<b>5.4</b>
Proposal	1	<b>1.6</b>	4	<b>5.3</b>	0	<b>0</b>	5	<b>2.5</b>
Literature Survey	1	<b>1.6</b>	0	<b>0</b>	0	<b>0</b>	1	<b>0.5</b>
	62	<b>100</b>	75	<b>100</b>	68	<b>100</b>	205	<b>100</b>

Progression in Engineering follows the building block approach described in Section 1 where specific skills, and then combinations of these skills, are targeted in separate assignments, culminating in long projects and design proposals where multiple skills are

integrated in assignments that reflect published academic writing and professional writing respectively. The demands on Engineers to develop a range of writing skills are therefore much greater than on Historians who have more opportunity to hone specific argumentation skills in a narrower range of genres.

## 7 DISCUSSION: BLENDING APPROACHES IN CORPUS DESCRIPTION

Adopting a research design which affords multiple perspectives on genres of assessed writing was a calculated decision. Rather than designing the project and corpus entirely to suit one methodology, we aimed to embrace different approaches. This brought constraints and benefits. The main constraints emerged in the process of designing the corpus itself as a digital corpus which would be freely available to other researchers. We successfully produced a significant corpus, the first of its kind of university student writing, and one whose design is now being emulated in the United States and elsewhere. This corpus is not fully aligned with any of our three descriptive methodologies, however. For instance, we have not fully investigated all the disciplinary contexts and do not have interviews with many of the students and tutors involved in producing the texts. Our final product is a compromise between corpus as text and corpus as texts, and as such has not been prepared exclusively with the multidimensional analysis or the genre analysis in mind.

This paper intends to point to the benefits of adopting multiple perspectives, and explores the extent to which this can be done through undergraduate History and Engineering. There is a stronger claim which could be made: that all four approaches are necessary for a satisfactory account of genres of assessed writing. The broad arguments for mutual benefit when ethnographic and linguistic approaches are combined are similar to those for linguistic ethnography, an approach that has gained in popularity in British applied linguistics circles (Creese 2008). They differ in specificity however. In the context of SFL, language is expected to be analysed in a context of situation and a context of culture (Eggins and Martin 1997), but the implication seems to be that linguists will be able to infer situation from language, and vice-versa. This works well for everyday, recognisable situations and cultures, but in the case of disciplinary discourses when researchers may well misconstrue meanings and inferences in terms of their own discourses, it would seem sensible to engage in explicit investigation of the target discourse communities and their practices.

The arguments for analysing genres in context are equally well rehearsed, both from Swalesean approaches which emphasise the role of discourse community and from SFL approaches which emphasise the importance of analysing language in its social context. Some studies of academic genres have explicitly referred to interviews with members of the discourse community, but other studies appear to assume that researchers can interpret texts in context without direct access to the understandings of those using them. Where Linguistic Ethnography tends to focus on the ethnography, with evidence from discourse analysis to develop critical interpretations, the genre approaches tend to focus on the texts, with evidence from their context to develop linguistic arguments. Somewhat worrying perhaps is the implication that genres can be identified before the context has been investigated. Veel and Coffin (1996:194-5) recognise the need for “the descriptions of language use ....[to be] much more

grounded in the particular contexts of school subjects than previous descriptions of genres” (194), and yet they still outline the protocol for analysing the literacy demands of secondary school subjects in the *Write It Right* project as starting with the analysis of genres, then linking this to context by locating the genres in relation to the syllabus, exams, programs, school assessment and classroom practice. As written, this is potentially circular, if the social purpose of the genre is determined prior to investigation of the context. Presumably there was interaction and modifications would result in ongoing description of register and lexicogrammar. Nevertheless, the research design as stated seems very much text first, context second, rather than the other way round. Our design supports a linguistic focus with explicit attention to the social context through interviews and document analysis. An ethnographically informed genre analysis surely increases the trustworthiness of the analysis.

Equally, there is increasing interest in two pronged (Matthiessen 2006) methods; in shunting between levels of the theory (Miller 2006, Halliday 2002); in shunting between different orders of magnitude, which Thompson and Hunston (2006:13) liken to an echo-chamber; and in shunting between the corpus and theory/description (Tucker 2006:87). In all these researchers can move back and forth between manual, interpretive accounts of individual texts to automated computerised analysis of large amounts of text. Where SFL approaches tend to start with theory driven insights, corpus linguists tend to start with patterns in text. Biber’s MDA provides an excellent example of patterns seeking explanation which raise future challenges for SFL theory. Indeed I would argue that without the contextual work on text types or SFL analysis into genre families, it would be difficult to make sense of the MDA results.

Our multidimensional research design is ambitious in its integration of approaches. Theoretical contradictions have not been explored, but to the extent that data has been compatible, it is because methods share assumptions such as an interest in real language use in context; an assumption that lexico-grammatical features cluster to shape register in ways that reflect the communicative context; a ‘theory of functional variation: of how texts are different, and the contextual motivations for those differences’ (Eggins & Martin 1997:236). I have assumed a broadly SFL theoretical perspective on genre, embracing the schematic structures developed by Martin (1992:105) and the notion of generic structure potential from Hasan (1996), but found it useful to extend theory by introducing a notion of ‘text type’ to refer to genres identified by the discourse communities, and the notion of ‘genre families’, a term used slightly differently by Martin (1992), to group similar genres across disciplines, which here is prior to full genre descriptions. Such proliferation of terminology might not be needed in single method approaches. Equally, there might have been fewer compromises in sampling strategy in a study from a single theoretical perspective.

Our multiple approaches have benefits in terms of bringing in different perspectives on the assignments. Indeed, they can strengthen findings from any given approach, and push investigators to probe contradictions or explain patterns begging interpretation. The genre analysis of one text is enhanced by understanding of its intended audience and purpose from the discourse community, by empirical evidence from automated corpus analysis, and from understanding in relation to other genres in

the same family or the same discipline. The corpus with its detailed tagging<sup>5</sup> of features and contextual metadata for each file, which includes information such as genre family, is now a rich marked up resource for future research. There is of course much more that can be said about History and Engineering assignments, and there are other studies that invite useful comparison (e.g. Byrd 2005; Cortes 2004; Ravelli 2004; Veel and Coffin 1996; Martin 2003; McCabe 2004; Ward 2007); so many further studies are envisaged.

Potentially our descriptions of genres of assessed writing will be richer, informed as they will be by a four dimensional approach. In this paper I have used illustrations from History and Engineering to suggest how a fuller description could encompass a rounded view of assignments from the discourse community, from the lexico-grammar and from the genre perspectives where the methods suggest contradictory findings, the benefits of multiple angles emerge as these contradictions are resolved and fuller understandings emerge. It is anticipated that more examples of this benefit will emerge as our description of genres of assessed writing continues.

#### REFERENCES

- Alsop, Sian and Hilary Nesi. 2008. Issues in the development of the British Academic Written English (BAWE) corpus. Unpublished paper under review for *Corpora*
- Becher, T. & P. Trowler. 2001. *Academic Tribes and Territories*. 2nd ed. Milton Keynes: Open University Press.
- Biber, Douglas 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Braxton, J. (1995) Disciplines with an affinity for the improvement of undergraduate education. In Hativa, N. and Marincovich, M. (eds) *Disciplinary differences in teaching and learning*. San Francisco: Jossey-Bass Publishers. 59--64
- Charles, Maggie 2006. The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied Linguistics* 27 (3): 492–518.
- Coffin, Caroline. 2004. Learning to write History: The role of causality. *Written Communication* 21, 261-289.
- Conrad, Susan. 2001. Variation among disciplinary texts: a comparison of textbooks and journal articles in biology and history. In Conrad and Biber (eds). Pp. 94-107. *Variation in English: multi-dimensional studies*. Harlow: Pearson.
- Conrad, Susan and Douglas Biber. 2001. Multi-dimensional methodology and the dimensions of register variation in English. In Conrad and Biber (eds) pp. 13-42. *Variation in English: multi-dimensional studies*. Harlow: Pearson.

---

<sup>5</sup> I would acknowledge here other members of the project team, particularly those who worked so hard on collection and markup which make the description of genres possible: Sian Alsop, Dawn Hindle, Maria Leedham, Signe Ebeling, Alois Heuboeck, Jasper Holmes, Richard Forsyth, Laura Powell, Paul Thompson, Paul Wickens and Hilary Nesi (PI).



- Creese, Angela. 2008. *Linguistic Ethnography*. In K.A. King and N.H. Hornberger (eds) *Encyclopedia of Language and Education*, 2<sup>nd</sup> Ed. Volume 10: *Research Methods in Language and Education*, 229-241. Springer.
- Ebeling, Signe & Alois Heuboeck. 2007. Encoding document information in a corpus of student writing: the experience of the British Academic Written English (BAWE) corpus. *Corpora* 2/2.
- Eggins, Suzanne and J.R. Martin. 1997. Genres and registers of discourse. In T.A. vanDyke (ed.) pp 230-256. *Discourse as Structure and Process*. London: Sage.
- Freedman, Aviva, Christine Adam and Graham Smart. 1994. Wearing suits to class: simulating genres and simulations as genre. *Written Communication*, 11(2), 193-226.
- Gardner, Sheena & Dawn Hindle. 2005. An emic perspective on assessed genres: gathering contextual evidence. In H. Nesi, S. Gardner, R. Forsyth, D. Hindle, P. Wickens, S. Ebeling, M. Leedham, P. Thompson, & A. Heuboeck 'Towards the compilation of a corpus of assessed student writing: An account of work in progress' in P. Danielsson & M. Wagenmakers (eds) *Corpus Linguistics* 1(1) Online at [www.corpus.bham.ac.uk/PCLC](http://www.corpus.bham.ac.uk/PCLC)
- Gardner, Sheena & Jasper Holmes. forthcoming Section headings and assignment macrostructures in undergraduate student writing. In E. Swain (ed.) *Thresholds and Potentialities of Systemic Functional Linguistics: Applications to other disciplines, specialised discourses and languages other than English*. Trieste: Edizioni Universitarie Trieste.
- Gardner, Sheena & Laura Powell. 2006. An investigation of genres of assessed writing in British higher education: a Warwick-Reading-Oxford Brookes project. Paper presented at seminar "Research, scholarship and practice in the area of Academic Literacies", University of Westminster, 30 June.
- Gardner, Sheena. in press. Initial Sentence Subjects and Angle on Field: Mapping ideational meaning across disciplines and years of study in a corpus of assessed student writing. In C. Jones and E. Ventola (eds) *Field, Ideation & Experiential Representation: from Language to Multimodality*. Series: Functional Linguistics edited by R. Fawcett. London: Equinox Publishing.
- Groom, Nicholas. 2005. Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes* 4, 257-277
- Halliday, M.A.K. 1961. Categories of the theory of grammar. *Word* 17:241-92. Reprinted in J.J. Webster (ed.) 2002. *On grammar, Vol. 1 in the Collected Works of M.A.K. Halliday*. Pp. 37-94. London: Continuum.
- Hasan, R. 1996. The nursery tale as genre. In C. Cloran, D. Butt and G. Williams (eds) *Ways of Saying, Ways of Meaning*. London: Cassell.
- Hyland, Ken 2004. Disciplinary interactions: metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*. 13, 133-151.
- Kelly, T., Richards, L. & Nesi H. 2004 Seminar Skills 1: Presentations
- Kelly, T., Sharpling, G. & Nesi H. 2006 Seminar Skills 2: Discussions
- Martin, J.R. 1992. *English Text: System and Structure* Amsterdam: Benjamins.
- Martin, J.R. 2003. Making history: grammar for interpretation. In J.R. Martin & R. Wodak. *Rereading the Past: critical and functional perspectives on time and value*. Amsterdam: John Benjamins.
- Martin, J.R. and David Rose. 2003. *Working with Discourse: meaning beyond the clause*. London: Continuum.

- Matthiessen, C.M.I.M. 2006. Frequency profiles of some basic grammatical systems: an interim report. In G. Thompson & S. Hunston (eds) *System and Corpus: exploring connections*. 103-142. London: Equinox.
- Miller, D. 2006. From concordance to text: appraising 'giving' in Alma Mater donation requests. In G. Thompson and S. Hunston (eds) *System and Corpus: exploring connections*. 248-268. London: Equinox.
- Nesi, Hilary & Sheena Gardner 2006 Variation in disciplinary culture: University tutors' views on assessed writing tasks. In: Kiely, R., Rea-Dickins, P., Woodfield, H. & Clibbon, G., (eds) *Language, Culture and Identity in Applied Linguistics. British Studies in Applied Linguistics* 21. London: Equinox Publishing. 99-117.
- Neumann, R. 2001 Disciplinary differences and university teaching. *Studies in Higher Education* 26: 135--146
- Neumann, R., Parry, S. and Becher, T. 2002 Teaching and learning in their disciplinary contexts: a conceptual analysis. *Studies in Higher Education* 27: 405--417
- Ravelli, L.J. 2004. Signalling the organization of written texts; hyper-Themes in management and history essays. In L.J. Ravelli & R.A. Ellis (eds) *Analysing Academic Writing: Contextual Frameworks*. London: Continuum.
- Samraj, Betty. 2005. An exploration of a genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes* 24, 2:141-156.
- Stubbs, Michael. 2006. Corpus Analysis: the state of the art and three types of unanswered questions. In G. Thompson & S. Hunston (eds) *System and Corpus: exploring connections*. 15-36. London: Equinox.
- Thompson, Geoff & Susan Hunston. 2006. System and Corpus: two traditions with a common ground. In G. Thompson & S. Hunston (eds) *System and Corpus: exploring connections*. 1-14. London: Equinox.
- Tucker, Gordon. 2006. Systemic incorporation: on the relationship between corpus and systemic functional grammar. In G. Thompson & S. Hunston (eds) *System and Corpus: exploring connections*. 81-102. London: Equinox.
- Veel, Robert & Caroline Coffin. 1996. Learning to think like an historian: the language of secondary school History. In R. Hasan and G. Williams (eds.) pp. 191-231. *Literacy in Society*. Harlow: Addison Wesley Longman.
- Ward, Jeremy. 2007. Collocation and technicality in EAP engineering. *Journal of English for Academic Purposes* 6, 18-35.
- Xiao, Zhonghua & Anthony McEnery. 2005. Two approaches to genre analysis: three genres in modern American English. *Journal of English Linguistics*, 33, 1, 62-82.

## APPENDIX

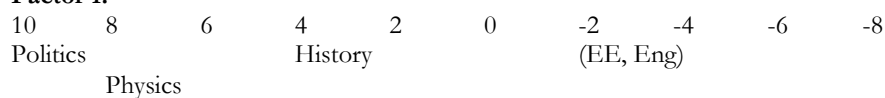
### THE BAWE FACTORS

Although the results for the MDA were not available at the 2007 conference, they are included here to suggest their potential. Four factors emerged as significant in the BAWE Corpus.

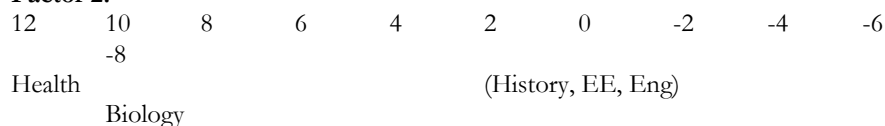
Factors 1, 2 and 3 all show significant differences between the Disciplinary Groups, with AH and PS at opposite ends of the scales for Factors 2 and 3.

Factors 1, 3 and 4 all show significant differences between History and Engineering. The interpretation of these academic writing factors is to be done, but the greater differentiation here, compared to the 1998 dimensions, and the plotting of 28 disciplines suggest new insights on academic writing should emerge.

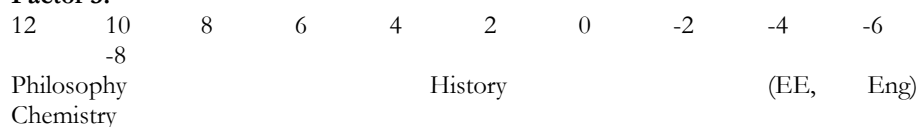
**Factor 1:**



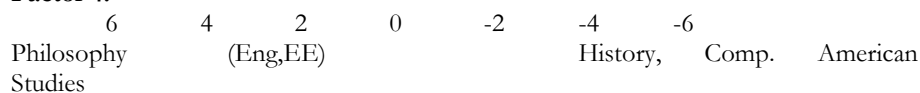
**Factor 2:**



**Factor 3:**



**Factor 4:**



Sheena Gardner  
School of Education  
University of Birmingham  
Edgbaston  
Birmingham  
B15 2TT

s.f.gardner@bham.ac.uk